

FINAL EXAM - CSC 294 - May 6, 2014

NAME: _____

ID: _____

Answer all 10 questions in the space provided. All questions have equal weight. Use the back of the page for additional space if needed (please indicate clearly where you have done this). Critical values for the standard normal, t , χ^2 and F distributions will be given as needed. Probabilities associated with any other distributions may be evaluated using normal approximation methods. A methodological summary is included at the end of the exam book.

Q1: For an *iid* sample from a normal distribution $N(\mu, \sigma^2)$ we are given sample mean $\bar{X} = 9.362$, $n = 56$, sample standard deviation $S = 4.912$.

- (a) Calculate a confidence interval for population mean μ with confidence level $1 - \alpha = 0.95$.
- (b) Calculate a confidence level $1 - \alpha = 0.95$ upper bound for σ .
- (c) Using the upper bound for σ calculated in part (b) estimate the sample size needed to obtain a confidence interval for μ with a margin of error of 0.75. Use confidence level $1 - \alpha = 0.95$.

$$[t_{55,0.025} = 2.00, \chi_{55,0.95}^2 = 38.96, z_{0.025} = 1.96]$$

Q1 [Answer]:

(a)

$$\begin{aligned} CI_{1-\alpha} &= \bar{X}_n \pm t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} \\ &= 9.362 \pm 2.00 \times \frac{4.912}{\sqrt{56}} \\ &= 9.362 \pm 1.315 = (8.047, 10.677) \end{aligned}$$

(b)

$$\begin{aligned} UB &= \frac{S}{\sqrt{\chi_{n-1, 1-\alpha}^2 / (n-1)}} \\ &= \frac{4.912}{\sqrt{38.96/55}} \\ &= 5.836 \end{aligned}$$

So,

$$\sigma < 5.836$$

is the 95% upper confidence bound for σ .

- (c) Use estimate $\hat{\sigma} = 5.836$ in formula

$$n \approx \left(z_{\alpha/2} \frac{\hat{\sigma}}{E_o} \right)^2 = \left(1.96 \times \frac{5.836}{0.75} \right)^2 = 232.6216,$$

so round up to $n = 233$.

Q2: We are given two independent samples from normally distributed populations. The sample means, sample standard deviations and sample sizes are summarized in the table below.

- (a) Use an F -test to test for equality of variances, using significance level $\alpha = 0.05$.
- (b) Using the appropriate procedure based on the test for equality of variances, calculate a confidence interval for $\mu_1 - \mu_2$ with confidence level $1 - \alpha = 0.99$.

	Sample 1	Sample 2
\bar{X}_i	44.673	44.299
S_i	0.782	0.923
n_i	35	36

$$[F_{34,35,0.975} = 0.506, F_{34,35,0.025} = 1.968, t_{69,0.005} = 2.649]$$

Q2 [Answer]:

- (a) Use statistic

$$F = \frac{S_1^2}{S_2^2} = \frac{0.782^2}{0.923^2} = 0.718.$$

Reject $H_0 : \sigma_1^2 = \sigma_2^2$ if

$$F \leq F_{1-\alpha/2, n_1-1, n_2-1} = 0.506 \text{ or } F \geq F_{\alpha/2, n_1-1, n_2-1} = 1.968$$

Therefore, do not reject the null hypothesis of equal variances at a significance level $\alpha = 0.05$ (P -value = 0.3361).

- (b) Use the pooled procedure with $\nu = n_1 + n_2 - 2 = 69$ degrees of freedom. Pooled variance is given by

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0.7335$$

The confidence intervals

$$\begin{aligned}
 CI_{1-\alpha} &= \bar{X}_2 - \bar{X}_1 \pm t_{n_1+n_2-2, \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\
 &= \bar{X}_2 - \bar{X}_1 \pm t_{69, 0.005} S_p \sqrt{\frac{1}{35} + \frac{1}{36}} \\
 &= 44.299 - 44.673 \pm 2.649 \sqrt{0.7335} \sqrt{\frac{1}{35} + \frac{1}{36}} \\
 &= 0.374 \pm 0.539 = (-0.913, 0.165).
 \end{aligned}$$

Q3: We are given two paired samples from normally distributed populations ($n = 6$). The data is summarized in the table below. Perform a two-sided hypothesis test for a difference in mean, using hypotheses $H_o : \mu_1 - \mu_2 = 0$ against $H_a : \mu_1 - \mu_2 \neq 0$. Use significance level $\alpha = 0.05$.

	Sample 1	Sample 2	Difference
1	11.709	10.699	1.010
2	11.894	10.488	1.406
3	11.427	10.342	1.085
4	11.971	10.558	1.413
5	11.818	10.903	0.915
6	12.348	10.871	1.477

$$[t_{5,0.025} = 2.571]$$

Q3 [Answer]:

$$\bar{X}_1 = 11.861, \bar{X}_2 = 10.643, S_D = 0.242.$$

Test statistic is

$$\begin{aligned}
 T &= \frac{\bar{D}}{S_D/\sqrt{n}} \\
 &= \frac{\bar{X}_1 - \bar{X}_2}{S_D/\sqrt{n}} \\
 &= \frac{1.218}{0.242/\sqrt{6}} \\
 &= 12.317.
 \end{aligned}$$

Reject H_o if

$$|T| \geq t_{n-1,\alpha/2} = t_{5,0.025} = 2.571.$$

Therefore, reject the null hypothesis at a significance level $\alpha = 0.05$ (P -value = 6.246e-05).

Q4: A model predicts that in a given population of trout, on average 13 out of every 40 specimens will possess a certain genetic mutation. Out of a random sample of size $n = 174$ the mutation is observed in $X = 51$ specimens.

- (a) Perform an appropriate two-sided hypothesis test to determine whether or not the data is compatible with the model. Use significance level $\alpha = 0.05$. Use a continuity correction.
- (b) Assuming that the actual population proportion with the mutation will not exceed $p = 0.4$, what sample size is needed to obtain a confidence interval for p with confidence level $1 - \alpha = 0.95$ and a margin of error no greater than $E = 0.05$?

$$[z_{0.025} = 1.96]$$

Q4 [Answer]:

- (a) We are testing

$$H_o : p = p_0 \text{ against } H_a : p \neq p_0$$

where $p_0 = 13/40 = 0.325$.

$$\hat{p} = 51/174 = 0.293,$$

With continuity correction, since $\hat{p} < p_0$ use

$$\begin{aligned} Z_{obs} &= \frac{X + 0.5 - np_0}{\sqrt{np_0(1 - p_0)}} \\ &= \frac{51 + 0.5 - 174 \times 0.325}{\sqrt{174 \times 0.325 \times (1 - 0.325)}} \\ &= -0.8174. \end{aligned}$$

(Without continuity correction $Z = -0.898$).

Reject H_o if

$$|Z| \geq z_{\alpha/2} = 1.96.$$

Therefore, so not reject the null hypothesis at a significance level $\alpha = 0.05$.

- (b) If we assume $p \leq 0.4$ then $p(1 - p)$ is maximized by substituting $p^* = 0.4$, giving conservative estimate

$$\begin{aligned} n &= p^*(1 - p^*) \left(\frac{z_{\alpha/2}}{E} \right)^2 \\ &= 0.4(1 - 0.4) \left(\frac{1.96}{0.05} \right)^2 \\ &= 368.7936. \end{aligned}$$

Rounding up gives sample size $n = 369$.

Q5: A study examined $n = 969$ male heart attack patients. Each subject was classified as having normal blood pressure (NBP) or high blood pressure (HBP). In each addition, each subject was observed for two years following the initial heart attack. Of 495 subjects with NBP 123 experienced a recurrent heart attack, while out of 474 subjects with HBP, 145 experienced a recurrent heart attack.

	NBP	HBP
Recurrence	123	145
No recurrence	372	329
n_i	495	474
\hat{p}_i	0.248	0.306

- (a) Construct a confidence interval for the difference in recurrence rates between the NBP and HBP subjects. Use confidence level $1 - \alpha = 0.95$.
- (b) Construct a confidence interval for the log odds ratio:

$$\log \left[\frac{\text{Odds}(\text{Recurrence} \mid \text{NBP})}{\text{Odds}(\text{Recurrence} \mid \text{HBP})} \right].$$

Use confidence level $1 - \alpha = 0.95$.

- (c) What can be concluded from parts (a) and (b) as to whether or not recurrence rates differ between the NBP and HBP groups? Are the conclusions from parts (a) and (b) consistent?

$$[z_{0.025} = 1.96]$$

Q5 [Answer]:

- (a) We have estimates $\hat{p}_1 = 0.248$, $\hat{p}_2 = 0.306$. The confidence interval is then

$$\begin{aligned} CI_{1-\alpha} &= \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= 0.248 - 0.306 \pm 1.96 \sqrt{\frac{0.248(1-0.248)}{495} + \frac{0.306(1-0.306)}{474}} \\ &= -0.0574 \pm 0.0563 = (-0.114, -0.00112). \end{aligned}$$

- (b) The estimate of the OR is

$$\hat{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = 123 \times 329 / (145 \times 372) = 0.750.$$

The standard error is

$$SE(\log(OR)) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} = \sqrt{\frac{1}{123} + \frac{1}{145} + \frac{1}{372} + \frac{1}{329}} = 0.1440636.$$

An approximate $(1 - \alpha)100\%$ confidence interval for $\log(OR)$ is therefore

$$\begin{aligned} CI_{1-\alpha} &= \log(\hat{OR}) \pm z_{\alpha/2} SE(\log(OR)) \\ &= \log(0.750) \pm 1.96 \times 0.1441 \\ &= -0.2874 \pm 0.2824 = (-0.5698, -0.005). \end{aligned}$$

- (c) From part (a), the CI for $p_1 - p_2$ is entirely negative. From part (b) the CI for OR is also entirely negative. Both conclusions imply $p_1 < p_2$, and so are consistent.

Q6: Four subspecies of a type of plant are hypothesized to occur in a given habitat according to the hypothetical relative frequencies $(p_1, p_2, p_3, p_4) = (0.45, 0.15, 0.15, 0.25)$. A random sample of $n = 260$ plants results in the observed counts for each subspecies:

	1	2	3	4	Totals
Observed counts O_i	133	57	51	19	260
Hypothetical frequencies p_i	0.45	0.15	0.15	0.25	1.00
Expected counts					260

- (a) For each subspecies, calculate the expected count assuming that the hypothetical frequencies are true. Place your answers in the table above. Does one subspecies appear to be overrepresented?
- (b) Use a χ^2 test for null hypothesis $H_o : p_i$ are the true population frequencies. Use significance level $\alpha = 0.05$. Yate's correction is not needed.

$$[\chi_{3,0.05}^2 = 7.815]$$

Q6 [Answer]:

- (a) According to the table, species 4 is *under-represented* in the sample.

	1	2	3	4	Totals
Observed counts O_i	133	57	51	19	260
Expected counts E_i	117.00	39.00	39.00	65.00	260
$(O_i - E_i)^2 / E_i$	2.19	8.31	3.69	32.55	46.74

- (b) Without Yate's correction: $X^2 = 46.742$. Reject H_o if

$$X^2 \geq \chi_{k-1, \alpha}^2 = 7.815.$$

Therefore, reject the null hypothesis at a significance level $\alpha = 0.05$ (P -value = 3.944e-10).

Q7: A contingency table with $n_r = 2$ rows and $n_c = 3$ columns based on a random sample of size $n = 555$ is given below. Hypothetical population frequencies of cell i, j are given by $p_{i,j}$. The population frequencies for the marginal row i and column j categories are given by r_i and c_j , respectively. Use a χ^2 test for the null hypothesis of row and column independence $H_o : p_{i,j} = r_i c_j$ for all i, j . Use significance level $\alpha = 0.05$. Yate's correction is not needed.

	1	2	3	Totals
1	162	72	78	312
2	77	82	84	243
Totals	239	154	162	555

Table 1: Observed counts $O_{i,j}$

$$[\chi^2_{2,0.05} = 5.991]$$

Q7 [Answer]: Without Yate's correction: $X^2 = 22.877$. Reject H_o if

$$X^2 \geq \chi^2_{(n_r-1)(n_c-1),\alpha} = 5.991.$$

Therefore, reject the null hypothesis at a significance level $\alpha = 0.05$ (P -value = 1.077e-05).

Expected counts $E_{i,j}$				
	1	2	3	Totals
1	134.00	87.00	91.00	312.00
2	105.00	67.00	71.00	243.00
Totals	239.00	154.00	162.00	555.00

X^2 statistic terms $(O_i - E_i)^2/E_i$				
	1	2	3	Totals
1	6.00	2.00	2.00	10.00
2	7.00	3.00	2.00	12.00
Totals	13.00	5.00	4.00	22.00

Q8: Independent samples for $k = 3$ treatments are summarized in the table below. Assume sample j is from a normally distributed population with mean μ_j and fixed variance σ^2 .

	1	2	3	4	5	\bar{X}_i	S_i	n_i
Treatment 1	13.13	15.16	10.60	16.41	17.99	14.66	2.88	5
Treatment 2	9.04	6.94	9.30	9.00	10.69	8.99	1.34	5
Treatment 3	20.55	16.37	20.20	14.39	21.89	18.68	3.16	5

(a) Construct an ANOVA table (fill in the 9 spaces in the ANOVA table below).

	SS	DF	MS	F
Treatment	_____	_____	_____	_____
Error	_____	_____	_____	
Total	_____	_____		

(b) Use an F -test for null hypothesis $H_o : \mu_i = \mu_j$ for all i, j . Use significance level $\alpha = 0.05$.

$$[F_{2,12,0.05} = 3.885]$$

Q8 [Answer]:

(a) ANOVA table is

ANOVA Table				
	SS	DF	MS	F
Treatment	236.83	2.00	118.42	17.70
Error	80.27	12.00	6.69	
Total	317.11	14.00		

(b) $F = 17.702$. Reject H_o if

$$F \geq F_{k-1, n-k, \alpha} = 3.885.$$

Therefore, reject the null hypothesis at a significance level $\alpha = 0.05$ (P -value = 0.0002631).

Q9: A new type of insecticide was tested against 3 standard alternatives. Each of the four insecticides was tested in 6 separate plots (requiring 24 separate plots). The percentage crop loss was recorded for each of the 24 plots at the end of the experiment. For each insecticide, the sample mean and sample standard deviation of the 6 outcomes is given in the following table (for example, the average percentage crop loss for the six plots using Standard Insecticide B was 9.91). The ANOVA table for the data is also given. Using a Bonferroni multiple comparison procedure, determine whether or not the new insecticide resulted in the lowest average percentage crop loss of all the insecticides tested. Use a familywise error rate of $\alpha_{FWE} = 0.05$.

	\bar{X}_i	S_i	n_i
New Insecticide	5.17	3.34	6
Standard Insecticide A	10.65	3.04	6
Standard Insecticide B	9.91	3.20	6
Standard Insecticide C	9.88	2.51	6

	SS	DF	MS	F
Treatment	113.69	3	37.90	4.10
Error	184.65	20	9.23	
Total	298.35	23		

$$[t_{20,0.05/6} = 2.613]$$

Q9 [Answer]:

We need $m = 3$ comparisons, to compare $\mu_1 - \mu_i$, $i = 2, 3, 4$. Since $n_i = 6$ for $i = 1, 2, 3, 4$, the CI s take form

$$\begin{aligned}
 CI &= \bar{X}_i - \bar{X}_j \pm t_{n-k, \alpha_{FWE}/(m2)} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \\
 &= \bar{X}_i - \bar{X}_j \pm t_{20, 0.05/6} \sqrt{9.23 \left(\frac{1}{6} + \frac{1}{6} \right)} \\
 &= \bar{X}_i - \bar{X}_j \pm 2.613 \sqrt{9.23 \left(\frac{1}{6} + \frac{1}{6} \right)} \\
 &= \bar{X}_i - \bar{X}_j \pm 4.58.
 \end{aligned}$$

The CI s are given in the following table. We can conclude with confidence $1 - \alpha_{FWE} = 0.95$ that μ_1 is the smallest mean, since $\mu_1 - \mu_i < 0$ for $i = 2, 3, 4$ within each comparison.

Multiple comparisons (Bonferroni procedure)						
	Treatment 1	Treatment 2	Difference	Margin of Error	LB	UB
Comp 1	1	2	-5.48	4.58	-10.06	-0.90
Comp 2	1	3	-4.73	4.58	-9.32	-0.15
Comp 3	1	4	-4.71	4.58	-9.30	-0.13

Q10: We are given two paired samples of sample size $n = 9$. The data is summarized in the table below. Suppose $\tilde{\mu}_D$ is the population median of the paired differences $D = X - Y$. Perform a two-sided signed rank test using hypotheses $H_o : \tilde{\mu}_D = 0$ against $H_a : \tilde{\mu}_D \neq 0$. Use significance level $\alpha = 0.05$, making use of the normal approximation. Make use of the table as needed.

	Sample 1 (X)	Sample 2 (Y)	Difference ($D = X - Y$)	Rank $ D $	Sign
1	5.4	4.9	0.5		
2	4.7	5.1	-0.4		
3	7.2	7.9	-0.7		
4	5.4	9.0	-3.6		
5	6.6	10.4	-3.8		
6	0.5	-3.0	3.5		
7	6.5	8.9	-2.4		
8	1.9	3.3	-1.4		
9	2.7	-0.3	3.0		

$$[z_{0.025} = 1.96]$$

Q10 [Answer]:

There are no ties, so $n' = 9$. The negative and positive rank sums are, respectively,

$$T_+ = 2 + 7 + 6 = 15 \text{ and } T_- = (10 \times 9)/2 - 15 = 45 - 15 = 30.$$

The mean and standard deviation of the negative or positive rank sums are

$$\mu_T = 22.5 \text{ and } \sigma_T = 8.441.$$

This gives

$$Z = \frac{T_+ - \mu_T}{\sigma_T} = \frac{15 - 22.5}{8.441} = -0.889.$$

Reject H_o if

$$|Z| \geq z_{\alpha/2} = 1.96.$$

Therefore, do not reject the null hypothesis at a significance level $\alpha = 0.05$ (P -value = 0.374).

	Sample 1 (X)	Sample 2 (Y)	Difference ($D = X - Y$)	Rank $ D $	Sign
1	5.4	4.9	0.5	2.0	+
2	4.7	5.1	-0.4	1.0	-
3	7.2	7.9	-0.7	3.0	-
4	5.4	9.0	-3.6	8.0	-
5	6.6	10.4	-3.8	9.0	-
6	0.5	-3.0	3.5	7.0	+
7	6.5	8.9	-2.4	5.0	-
8	1.9	3.3	-1.4	4.0	-
9	2.7	-0.3	3.0	6.0	+